

DOCUMENT RESUME

ED 160 921

CG 012 856

AUTHOR Kissler, Gary D.
TITLE Predicting Peer Evaluation of Scientists.
PUB DATE [77]
NOTE 20p.
AVAILABLE FROM Dr. G. D. Kissler, Navy Personnel Research and Development Center (Code 302), San Diego, CA 92152

EDRS PRICE MF-\$0.83 HC-\$1.67 Plus Postage.
DESCRIPTORS *Evaluation Methods; Government Employees; Job Skills; *Merit Rating Programs; *Peer Evaluation; *Personnel Evaluation; *Promotion (Occupational); Research Projects; Rewards; *Scientists

ABSTRACT

A sample of 103 government scientists was evaluated by panels of peers in regard to personnel action such as promotion. Analysis of information used by the panels and other data external to the process showed that peer ratings could be predicted accurately using only four variables: number of publications, honors and awards received, special assignments, and age. Six separate peer groups were compared across predicted peer ratings based on the cross-validated regression results above and no significant mean differences were found. A multivariate profile for promoted scientists suggested they were more recently educated, were lower in grade, had less time in grade, and had received more honors and awards than scientists who remained in grade. Some relationships between peer-ratings and performance-related measures were found to parallel similar results found by earlier researchers. Strengths and weaknesses of the peer rating system are discussed. (Author)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED160921

Predicting Peer Evaluation of Scientists

Gary D. Kissler

Navy Personnel Research and Development Center¹

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

G. D. Kissler

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Running head: Predicting Peer Evaluation

CG012856

Abstract

A sample of 103 Government scientists was evaluated by panels of peers for personnel action purposes. Analysis of information used by the panels and other data external to the process showed that peer ratings could be predicted accurately ($R = .84$; $p < .001$) using only four variables: number of publications, honors and awards received, special assignments, and age. Six separate peer groups were compared across predicted peer ratings based on the cross-validated regression results above and no significant mean differences were found. A multivariate profile for promoted scientists suggested they were more recently educated, were lower in grade, had less time in grade, and had received more honors and awards than scientists who remained in grade. Some relationships between peer-ratings and performance-related measures were found to parallel similar results found by earlier researchers. Strengths and weaknesses of the peer rating system are discussed.

Predicting Peer Evaluation of Scientists

A peer-evaluation process for research scientists was examined recently (Kissler, Note 3) and was found to compare favorably, with regard to its reliability and validity, to similar approaches reported in other organizational settings (Korman, 1968; Lewin & Zwany, 1976). It was stated that the relative predictive influence of data components in this process upon consensus peer-ratings had yet to be established. Further, earlier research has identified some characteristics of peer-ratings among non-research personnel which may apply to research scientists as well. Specifically, Weitz (1958) found peer evaluations to be significantly related to performance and Wherry and Fryer (1949) discovered peer ratings to be more highly related to performance than are supervisory ratings. Roadman (1964) was able to demonstrate that peer ratings can identify individuals who have been promoted more rapidly than others.

The purposes of this paper are (1) to examine the overall relationship between peer-ratings by scientists and variables concerning performance, position, and background characteristics of research incumbents; (2) to evaluate the independence of peer-evaluations for scientists from different peer groups; and (3) to compare across pertinent variables researchers who were or were not promoted, to suggest a profile of promoted scientists.

Method²Subjects

A group of 103 agricultural scientists was evaluated by peer-evaluation panels for purposes of promotional classification according to the Civil Service Commission Research Grade-Evaluation Guide (RGE) (CSC, Note 1).

For this analysis, the sample was categorized into six peer groups based on shared research subject matter, projects, and/or activities: (1) Animal ($\underline{n} = 19$), (2) Food/Chemistry/Nutrition ($\underline{n} = 20$), (3) Entomology ($\underline{n} = 18$), (4) Plant ($\underline{n} = 28$), (5) Environmental processing/mechanical engineering ($\underline{n} = 9$), and (6) Soil science ($\underline{n} = 7$). The sample peer group proportionately represented those in the larger government agency group of nearly 750 scientists from which they were drawn.

Measures

Peer ratings. Panel members of a Research Personnel Evaluation Panel (RPEP) reached unanimous consensus on scores across four factors, each said to represent "a different aspect of the job-incumbent relationship" (CSC, Note 1, p. 12). These factors and their respective components are given below (Handbook, Note 2):

Factor I: Research Assignment

1. Assigned responsibility
2. Research objective
3. Expected results

Factor II: Supervision Received

1. Assigned authority
2. Research guidance given
3. Review of results
4. General supervision

Factor III: Guidelines and Originality

1. Available literature
2. Originality required
3. Demonstrated originality

Factor IV: Qualifications and Contributions

1. Demonstrated research accomplishments
2. Research stature, recognition and impact on science and technology
3. Advisory and consultant activities

The stability coefficients between the previous and most recent peer-ratings for each were: I = .90, II = .90, III = .37, and IV = .92. Pre-consensus inter-rater reliabilities across these factors averaged 0.96 but the factor intercorrelations ranged from 0.89 to 0.96. Therefore, a single evaluation variable based on the total consensus score was used in this study.

Incumbent/position variables. Each scientist being evaluated provides the RPEP with information pertaining to the following variables: education, additional training, honors and awards, special invited addresses, memberships in professional societies, offices and committee assignments held in professional and honorary societies, participation in national scientific meetings, technical conferences, workshops, etc., professional advisory and consulting activities, special assignments of a technical and professional nature, and contributions to special foreign agricultural programs. The panel members are also informed of the incumbent's grade, tenure, and year of last academic degree earned. Age, although not explicitly stated, is generally inferred from information regarding education and past research experience.

The RPEP also receives a list of the incumbent's publications with those published since the last promotion so noted. For purposes of analysis aimed beyond quantity information, the publications were categorized as follows: refereed, nonrefereed, senior-author, junior-author, total

publications minus abstracts, and total publications.

Supervisory ratings. Each scientist receives a yearly supervisory performance evaluation which measures the categories quantity of work, quality of work, and supervision of others. Ratings in each category can range from 1 (unsatisfactory) to 9 (distinguished). These ratings were nearly always significantly correlated with peer evaluations and showed considerable stability across a 3-year period ($\bar{r} = .85$).

Results

A two-thirds random sample ($n = 77$) was drawn from the total group to generate initial regression results from predicting RPEP consensus scores. A preliminary examination of zero-order correlations between RPEP scores and various publication variables indicated three of these variables were most highly related to peer-ratings: number of refereed publications, total number of publications, and total publications excluding abstracts. Each of these variables was independently included with incumbent/position variables mentioned earlier to generate regression equations. Stepwise regression analysis was used to identify the fewest possible predictors that would yield a significant multiple correlation based on significant b -weights. The coefficients from each of three equations were used to obtain three sets of predicted RPEP scores for a hold-out sample ($n = 27$). The multiple regression coefficients based on the initial sample results were compared with corresponding cross-validated multiple coefficients which resulted from correlating the predicted and actual RPEP scores for the holdout sample. The most parsimonious equation in terms of predictive accuracy and number of predictors will be discussed here. Table 1 contains the results of this approach.

Insert Table 1 About Here

The variables in Table 1 are listed in the order in which they entered the stepwise equations. The sequence of predictors, ranging from high to low predictive efficiency, was: Total Publications, Honors and Awards, Special Assignments, and Age, yielding multiple and cross-validated coefficients of 0.84 ($p < .001$) and 0.78 ($p < .01$), respectively.

Based on the predictors in the equation, the total sample was used to obtain regression coefficients for creating a predicted RPEP score. Earlier research (Kissler, Note 3) had shown that peer groups received independent reviews by the RPEP. However, this finding was based on a cross tabulation of RPEP decisions (promote vs. remain in grade) and peer group membership. A more sensitive peer groups comparison was made by using the predicted RPEP variable in an analysis of variance design. The mean predicted RPEP score for each peer group was simultaneously compared with means for the remaining groups. The resulting F -ratio (1.99 with 5, 93 df) did not reach a conventional level of statistical significance.

To identify a possible "profile" pertaining to promoted versus remain-in-grade scientists, it was necessary to compare these groups across all incumbent/position variables simultaneously. This simultaneous comparison ruled out the use of t -tests to avoid taking advantage of chance differences. With this consideration and the possibility of interrelated independent variables in mind, Hotelling's T -squared approach (Hotelling, 1931) was selected as an appropriate analytic tool. With a dichotomous dependent variable (1 = promote; 2 = remain in grade), it was possible to use

discriminant analysis to obtain the appropriate evaluation. Table 2 contains the results from the simultaneous test of these two groups across all incumbent/position variables.

Insert Table 2 About Here

The results given in Table 2 first indicate that significant differences among these variables exist between the two groups ($F = 3.88$; $p < .0001$). Further, it is important to compare the independent and partial F -ratios for each variable considered. Time in Grade, Degree Year, and Grade are the only variables which have significant F -ratios in both columns. Age, Tenure, Education, and Date of Last Promotion had initial significant independent F -ratios but their partial F -ratios are not significant. Perhaps somewhat more interesting are the variables which, when the effects of other variables are ignored, were not found to be significantly different when related to the RPEP promotion/classification decision, but became significant when their effects were adjusted for the influence of other variables. The list of variables meeting this criterion were: Total Publications, Honors and Awards, and Special Invitations. Two of these variables were the strongest predictors in the regression equation predicting total RPEP score. Of all the variables considered, Total Publications had the largest partial F -ratio (17.62 ; $p < .001$). Further, when Total Publications was interchanged with other publication variables, the analysis correctly classified persons into the two categories 84.5% of the time, a larger percentage than was found in any of the other analyses. The sign of the discriminate coefficient associated with each significant partial F -ratio allows one to determine

the direction of the difference between the two groups (e.g., promoted scientists earned their degrees more recently)

To investigate the relationships between peer-ratings and work-related variables which were found in earlier studies, the correlations in Table 3 were assembled. The performance variables related to publications were chosen for this analysis because of their pattern of significant relationships with ratings measures (Kissler, Note 3).

Insert Table 3 About Here

The first row in Table 3 contains the correlations between the RPEP total consensus scores and several performance variables: Number of Abstracts, Refereed Publications, Nonrefereed Publications, Senior-author Publications, Junior-author Publications, Total Publications minus abstracts, and Total Publications. All the coefficients in this row are statistically significant. Each supervisory rating coefficient was tested for a significant difference from the corresponding RPEP coefficient. With the exception of the Number of Abstracts variable, nearly all of the supervisory rating coefficients were significantly different from their RPEP counterparts. Also, supervisory rating coefficients were smaller than the RPEP coefficients in every instance.

Finally, to identify scientists who have advanced rapidly, a group was formed on the basis of tenure and age so that those scientists falling below the median years of tenure (16 years) and above the median grade (13) were categorized as having been rapidly promoted. This group was compared with the remaining scientists with a t-test of their RPEP total consensus

score means. The mean of the rapidly promoted group ($\bar{X} = 47.0$; $SD = 4.7$; $n = 12$) was significantly larger ($t = 6.67$; $p < .0001$) than the remaining group mean ($\bar{X} = 33.7$; $SD = 13.9$; $n = 91$). Though not reported here, results were identical when these groups were compared across all four RGEF factors.

Discussion

The current research was able to identify those variables which represent a central core to the agreement reached by peers who evaluate other researchers. Though presented with a fairly large number of variables related to research performance, the RPEP total consensus score was accurately predicted using only four of these variables: Total Publications, Honors and Awards, Special Assignments, and Age. Of these, the single best predictor was Total Publications and, by comparing regression results with interchanged publication variables, it was found that better prediction was obtained with a publication variable which emphasized quantity over quality (e.g., refereed publications).³ On the other hand, except for age, the remaining predictors may reflect the panel's tendency to consider a sort of compensating approach to evaluation; credit appears to be given for research products other than publications. With regard to age, it has been found earlier (Shaw, 1967) that older scientists in this organization tend to increase their publication rate over time. It seems logical that this same relationship exists with other related accomplishments as well.

It was encouraging to find that the peer groups did not differ statistically when compared on a regression-based predicted RPEP total consensus score. These results supported earlier findings stemming from a less sensitive analysis (Kissler, Note 3).

A "profile" of scientists who were promoted indicated that these individuals were more recently educated, were lower in grade, averaged less time in grade, had received fewer special invitations, had more publications, and received more honors and awards than researchers who remained in grade. These groups did not differ with regard to age, tenure, education, date of last promotion, training, number of research subordinates, foreign currency program contributions, memberships in professional societies, advising/consulting duties, special assignments, or scientific meetings attended. It appears that the peer-evaluation method used for research scientists has construct validity as evidenced by the findings which parallel those discovered by earlier research in the general area of peer-ratings. Specifically, peer evaluations not only were significantly related to performance (Weitz, 1958), they had a stronger relationship to performance than did supervisory ratings (Wherry & Fryer, 1949). The RPEP ratings were also significantly higher for persons who have been more rapidly promoted than others (Roadman, 1964).

What seems to be emerging from this and previous research is a picture of an evaluation process which shares the same strengths and weaknesses as when used to evaluate nonresearch personnel in other organizational settings. Its strength lies in its ability to generate consistent and stable evaluations across peer groups having diverse backgrounds, interests, and research programs. Further credit must be given on the basis of validity coefficients found earlier between these ratings and external measures of research performance and peer recognition (Kissler, Note 3). The weaknesses lie partly in the overlap of evaluation factors and the failure to recognize that only a few of the numerous variables

examined by RPEP panels actually serve as a basis for discriminating between scientists who do or do not get recommended for promotion.

In summary, this evaluation process currently employed by several Government research and development organizations offers a method by which scientists and their contributions can be judged reliably and with measurable validity. Despite some reservations concerning the emphasis placed on certain predictors over others, it is felt this system has much to recommend it.

Table 1
Stepwise Regression Results from Predicting
Peer-Evaluation Consensus Scores

Variable	<u>b</u> -weight	Partial- <u>F</u>
Total Publications	.25(.37) ^a	26.15***
Honors and Awards	.90(.34)	22.05***
Special Assignments	1.37(.34)	23.32***
Age	.33(.21)	9.02***
Multiple <u>R</u>	.84***	
Cross-validated <u>R</u>	.78**	

Note. The constant for the equation is 4.06.

^aBeta weights are shown in parentheses.

**p < .01

***p < .001

Table 2
Hotelling's T-Squared Comparison of Promoted and
Remain-in-Grade Scientists Across Incumbent/Position Variables

Variable	Independent F-Ratio	Partial F-Ratio	Discriminant Coefficient
Degree Year	13.67***	2.20*	-.47
Age	10.78**	1.52	-.40
Tenure	6.74*	.29	.12
Grade	6.44*	14.70***	1.06
Education	5.70*	1.76	-.25
Time in Grade	5.18*	2.57**	.30
Last Promotion	4.52*	— ^a	
Training	3.75	.83	.15
Number of Subordinates	1.69	— ^a	
Special Invitations	1.03	2.15*	.28
Total Publications	1.01	17.62***	-.84
Foreign Program Contributions	.63	.14	-.06
Honors and Awards	.54	4.44***	-.38
Professional Societies	.06	.75	-.12
Advising/Consulting	.05	.47	-.12
Special Assignments	.04	1.70	-.19
Scientific Meetings	.04	1.21	-.18

Note. Wilks Lambda, $F = 3.88$; $p < .001$; $df = 15, 68$.

^aVariable not of sufficient strength to enter analysis

* $p < .05$

** $p < .01$

*** $p < .001$

Table 3

Correlations Between Scientific Performance Variables and Peer- and Supervisor-Rating

	Abstracts	Refereed Publications	Nonrefereed Publications	Senior- Author Publications	Junior- Author Publications	Total Publications (w/c Abstracts)	Total Publications
RPEP Score	.21*	.54***	.47***	.47***	.54***	.60***	.62***
Quantity	.10	.36***	.14 ^a	.20* ^a	.33* ^a	.31*** ^a	.32*** ^b
Quality	.13	.29** ^a	.20* ^a	.18* ^a	.29** ^a	.28** ^b	.30** ^b
Supervision	.06	.20* ^b	.09 ^b	.12 ^a	.18 ^b	.17 ^c	.18 ^c

Note. n ranged from 102-103 for RPEP coefficients and from 71-89 for supervisory rating coefficients.

W = weighted; U = Unweighted.

* $p < .05$

** $p < .01$

*** $p < .001$

^aSignificantly different from the corresponding RPEP coefficient at $p < .05$

^bSignificantly different from the corresponding RPEP coefficient at $p < .01$

^cSignificantly different from the corresponding RPEP coefficient at $p < .001$

Reference Notes

1. CSC research grade-evaluation guide (TD 52), June 1964.
2. Handbook for using the research grade-evaluation guide in ARS.
ARS Personnel Division, Classification and Pay Branch, Hyattsville,
Maryland (20782).
3. Kissler, G. D., Evaluating research activity: A scientific peer-rating
approach. (Working Paper, 1977).

References

- Hotelling, H. The generalization of student's ratio. Annals of Mathematical Statistics, 1931, 2, 360-378.
- Lewin, A. Y., & Zwany, A. Peer nominations: A model, literature critique and a paradigm for research. Personnel Psychology, 1976, 29, 423-447.
- Roadman, H. E. An industrial use of peer ratings. Journal of Applied Psychology, 1964, 48, 211-214.
- Shaw, B. T. The use of quality and quantity of publication as criteria for evaluating scientists. Miscellaneous Publication No. 1041, Agricultural Research, USDA, January 1967.
- Weitz, J. Selecting supervisors with peer ratings. Personnel Psychology 1958, 11, 25-35.
- Wherry, R. J., & Fryer, D. C. Buddy ratings: Popularity contest or leadership criterion? Personnel Psychology, 1948, 2, 147-159.

Footnotes

¹This study was conducted while the author was employed by the Agricultural Research Service, USDA. Requests for reprints should be sent to Dr. G. D. Kissler, Navy Personnel Research and Development Center, (Code 302), San Diego, CA 92152.

²The characteristics of subjects and measures in this study are extensively reviewed in an earlier paper (Kissler, Note 3). Tables of descriptive statistics related to variables included in this study are available upon request.

³It should be noted that refereed publications had a higher zero-order relationship to peer ratings and supervisor ratings than did nonrefereed publications. However, regression results using this publication variable required more predictors to achieve a multiple correlation coefficient not significantly different from the figure in Table 2.